

Experts ou (foule de) non-experts ?
la question de l'expertise des annotateurs
vue de la myriadisation (*crowdsourcing*)

Karën Fort

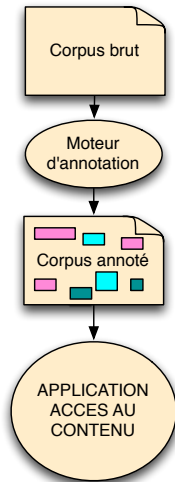
karen.fort@paris-sorbonne.fr



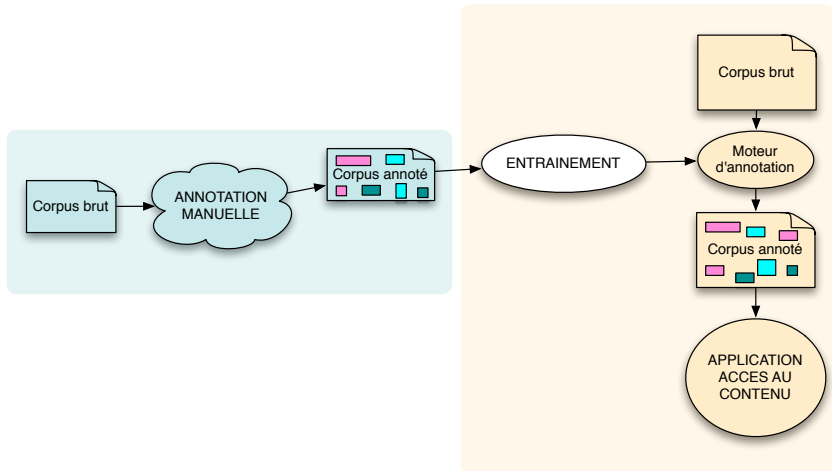
3 septembre 2015



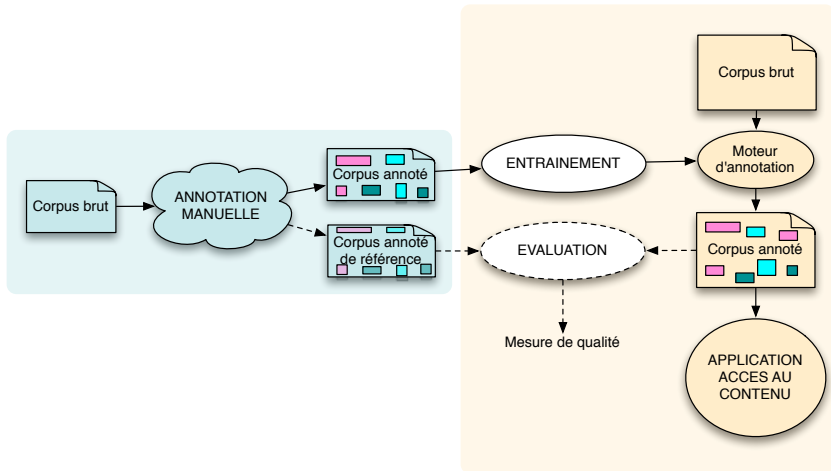
Corpus annotés et Traitement Automatique des Langues



Corpus annotés et Traitement Automatique des Langues



Corpus annotés et Traitement Automatique des Langues



Produire des corpus annotés manuellement : l'enjeu du coût

Prague Dependency Treebank [Böhmová et al., 2001] :

- ▶ 1,8 millions de mots annotés en morpho-syntaxe et syntaxe
- ⇒ 5 ans, 22 personnes (max. 17 en parallèle), 600 000 dollars

GENIA [Kim et al., 2008] :

- ▶ 9 372 phrases annotés en microbiologie (noms de gènes, protéines)
- ⇒ 5 annotateurs à temps partiel, 1 coordinateur senior et 1 junior pendant 1,5 ans

Que sait-on de l'annotation manuelle ?

Un existant parcellaire et mal documenté :

- ▶ bonnes pratiques de **haut niveau**
- ▶ solutions au **cas par cas** : biais ? qualité ?
- ▶ **bribes** de méthodologies
- ▶ mesures d'évaluation **pas toujours adaptées** et peu explicitées



Que sait-on de l'annotation manuelle ?

Un existant parcellaire et mal documenté :

- ▶ bonnes pratiques de **haut niveau**
- ▶ solutions au **cas par cas** : biais ? qualité ?
- ▶ **bribes** de méthodologies
- ▶ mesures d'évaluation **pas toujours adaptées** et peu explicitées

Nécessité d'une vision d'ensemble :

- **formaliser** l'annotation [Fort et al., 2012]
- **outiller** l'annotation [Fort and Sagot, 2010]
- définir les **mesures d'évaluation** adaptées [Mathet et al., 2012]
- clarifier le **rôle des acteurs**



L'annotation manuelle de corpus, un domaine en friches

La question de l'expertise des annotateurs

La myriadisation, un point de vue unique sur l'annotation

Conclusion et perspectives

L'expert, le naïf et le non-non

Le bon, la brute et le truand de l'annotation

5.3. Quels annotateurs (naïfs ou non, experts ou non) ?

[Tellier, 2014]

Par ailleurs, de nombreux textes ou extraits ont été multi-annotés par des **annotateurs naïfs et/ou experts** : plus de la moitié des textes annotés en relations rhétoriques sont annotés par deux annotateurs naïfs⁹ puis arbitrés par un expert¹⁰ (voir section 3.2).

9. Étudiants de linguistique de niveau licence 3, sans connaissances préalables des fondements théoriques du modèle d'annotation.

10. Chercheur participant au projet ANNODIS.

[Péry-Woodley et al., 2011]

Qu'est-ce qu'un expert ?

Pour les SCA, la durée de la perfusion dépend de la manière dont le SCA doit être traité : elle peut durer jusqu'à 72 heures au maximum chez les patients devant recevoir des médicaments.

Corpus *Sequoia* [Candito and Seddah, 2012], annoté en syntaxe :

- ▶ le sous-corpus (EMEA) est du domaine de la **pharmacologie**
- ▶ l'annotation est d'un certain type **linguistique** (syntaxe)

Qu'est-ce qu'un expert ?

Pour les SCA, la durée de la perfusion dépend de la manière dont le SCA doit être traité : elle peut durer jusqu'à 72 heures au maximum chez les patients devant recevoir des médicaments.

Corpus *Sequoia* [Candito and Seddah, 2012], annoté en syntaxe :

- ▶ le sous-corpus (EMEA) est du domaine de la **pharmacologie**
- ▶ l'annotation est d'un certain type **linguistique** (syntaxe)

⇒ qu'est-ce qu'un expert, dans ce cas ?

- ▶ un pharmacien ?
 - ▶ un linguiste ?
- ▶ un linguiste pharmacien ?

Et pourquoi pas ?

un locuteur du français,
«sans» connaissances en syntaxe ou en pharmacologie,
mais **formé** à la tâche ?

Définition

Dans le TLFi, *via* le CNRTL

■ **EXPERT, ERTE**, adj. et subst. masc.

I.- *Adjectif*

A.- **Expert à, en, dans...**[En parlant d'une pers., de ses gestes, de son activité]

1. Qui a acquis une grande habileté, un grand savoir-faire dans une profession, une discipline, grâce à une longue expérience. *En gars expert au maniement du bétail, il s'accroupit, tenta de lui saisir le pied droit avec sa main gauche et de le placer entre ses genoux* (ALAIN-FOURNIER, *Meaulnes*, 1913, p. 58).

Distinguer l'expertise :

1. du domaine du **corpus** (ici, la pharmacologie),
2. du domaine de l'**annotation** (ici, un certain type de syntaxe),
3. de la **tâche** (ici, annoter des relations syntaxiques avec tel ou tel outil, selon tel guide d'annotation)

L'annotation manuelle de corpus, un domaine en friches

La question de l'expertise des annotateurs

La myriadisation, un point de vue unique sur l'annotation

Conclusion et perspectives

La myriadisation (*crowdsourcing*)

[Sagot et al., 2011] ou *externalisation ouverte*, ou *production participative*

***Crowdsourcing** is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."*

[Howe, 2006]

- ▶ **pas d'identification** ou de sélection *a priori* des participants ("open call")

La myriadisation (*crowdsourcing*)

[Sagot et al., 2011] ou *externalisation ouverte*, ou *production participative*

Crowdsourcing is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."

[Howe, 2006]

- ▶ **pas d'identification** ou de sélection *a priori* des participants ("open call")
- ▶ **massive** (dans la production et la participation)

La myriadisation (*crowdsourcing*)

[Sagot et al., 2011] ou *externalisation ouverte*, ou *production participative*

Crowdsourcing is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."

[Howe, 2006]

- ▶ **pas d'identification** ou de sélection *a priori* des participants ("open call")
- ▶ **massive** (dans la production et la participation)
- ▶ (relativement) **peu coûteux** (répond à la préoccupation du coût)

Une taxinomie (simplifiée) de la myriadisation

parmi d'autres, voir [Geiger et al., 2011]

Wikipedia, Projet Gutenberg :

- ▶ bénévole (non rémunérée)
- ▶ directe (le but à atteindre n'est pas masqué)

Games With A Purpose (Jeux ayant un but ou GWAP) :

- ▶ bénévole (non rémunérée)
- ▶ indirecte (le but à atteindre est plus ou moins masqué)

Amazon Mechanical Turk :

- ▶ rémunérée
- ▶ directe

Une taxinomie (simplifiée) de la myriadisation

parmi d'autres, voir [Geiger et al., 2011]

Wikipedia, Projet Gutenberg :

- ▶ bénévole (non rémunérée)
- ▶ directe (le but à atteindre n'est pas masqué)

Games With A Purpose (Jeux ayant un but ou GWAP) :

- ▶ bénévole (non rémunérée)
- ▶ indirecte (le but à atteindre est plus ou moins masqué)

Amazon Mechanical Turk :

- ▶ rémunérée
- ▶ directe

Phrase Detectives : annotation d'anaphores

[Chamberlain et al., 2008]

Corpus annoté de 200 000 mots :

- ▶ corpus pré-annoté
- ▶ instructions détaillées
- ▶ phase de formation
- ▶ 2 manières de jouer

DETECTIVES CONFERENCE

Another detective has made a decision about a phrase, either that it refers to another phrase, it has not been mentioned before, it is a property or it does not refer to anything. Do you agree with them?

Knitta (Wikipedia)

PhyCell and Molyk came up with their own names, then invented names for other members in a brainstorming session they considered "one of the more hilarious meetings". Some former member names include Klotentous N.T., SorOfaBleth and P-Kraby.

As of January 2009, the group has two female members and one male, ages 30 to 73, who wish to remain anonymous. Current members are PhyCell, MaaucKoloby, and Grany SQ. An estimated five to twelve cryptic groups exist around the world.

Usually tagging on Friday nights and Sunday mornings, **knitta taggers** leave a paper tag on each work, bearing the slogan "knitta please" or "whadup knitta?". They tag trees, lamp posts, railings, fire hydrants, monuments and other urban targets, and even get a little "hardcore" with ideas like hanging knitted-bagged sweaters over aerial telephone cables. The crew marks holidays by doing themed work, using, for example, pink yarn for their Valentine's Day pieces and sparkly yarn for New Years. When Knitta is not working with a theme, they work on projects, tagging specific targets or specific areas.

The phrase and their followers consider their graffiti "a method of beautifying public space".

The phrase in blue is the **closest** phrase that refers to the phrase in orange.

Disagree Agree

NAME THE CULPRIT

Has the phrase shown in orange been mentioned before in this text or is it a property? Use your mouse to select the **closest phrase(s)** if it has been mentioned before.

Knitta (Wikipedia)

PhyCell and Molyk came up with their own names, then invented names for other members in a brainstorming session they considered "one of the more hilarious meetings". Some former member names include Klotentous N.T., SorOfaBleth and P-Kraby.

As of January 2009, the group has two female members and one male, ages 30 to 73, who wish to remain anonymous. Current members are PhyCell, MaaucKoloby, and Grany SQ. An estimated five to twelve cryptic groups exist around the world.

Usually tagging on Friday nights and Sunday mornings, Knitta taggers leave a paper tag on each work, bearing the slogan "knitta please" or "whadup knitta?". They tag trees, lamp posts, railings, fire hydrants, monuments and other urban targets, and even get a little "hardcore" with ideas like hanging knitted-bagged sweaters over aerial telephone cables. The crew marks holidays by doing themed work, using, for example, pink yarn for their Valentine's Day pieces and sparkly yarn for New Years. When Knitta is not working with a theme, they work on projects, tagging specific targets or specific areas.

knitta, they work on projects, tagging specific targets or specific areas.

Not mentioned before This is a property

Done

Phrase Detectives : évaluation des annotations produites [Chamberlain et al., 2009]

Existence d'une référence annotée par des "experts" :

- ▶ environ 94 % d'accord observé entre experts (référence)
- ▶ environ 84 % d'accord observé entre les joueurs et les experts

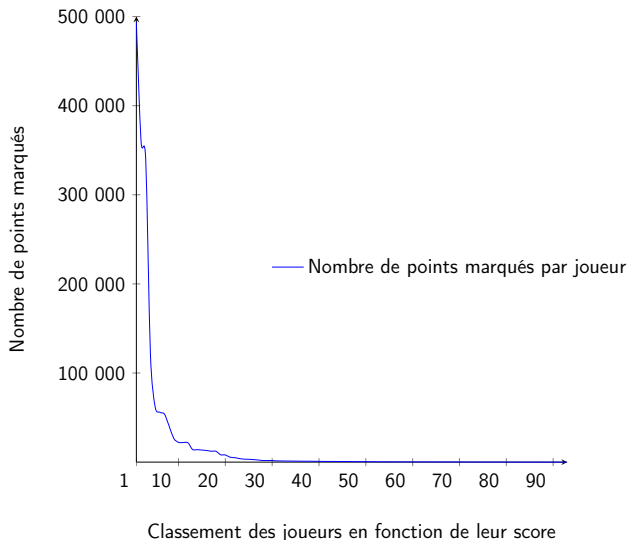
Échec (accord pratiquement nul) de l'annotation des propriétés
(*postman* est une propriété de *Jon*) :

Jon, the **postman**, delivered the letter

- + des joueurs sont devenus des experts de la tâche d'annotation en anaphore
- aucun n'est parvenu à maîtriser celle d'annotation des propriétés

Phrase Detectives : scores (production) des joueurs

Fév. 2011 - fév. 2012 (2 000 joueurs en tout)



ZombiLingo : annotation en syntaxe de dépendances [Fort et al., 2014]

ZOMBILINGO

Jouer Boutique Forum Joueurs

Prochain niveau : 10%

Phénomène : Sujet

Besoin d'aide?

Pour les SCA, la durée de la perfusion **DÉPEND** de la manière dont le SCA doit être traité : elle peut durer jusqu' à 72 heures au maximum chez les patients devant recevoir des médicaments.

48 15

Accent mis sur la formation et l'évaluation :

- ▶ décomposition de la tâche par phénomènes
- ▶ tutoriel par phénomène
- ▶ phrases de référence proposées régulièrement

ZombiLingo : annotations produites

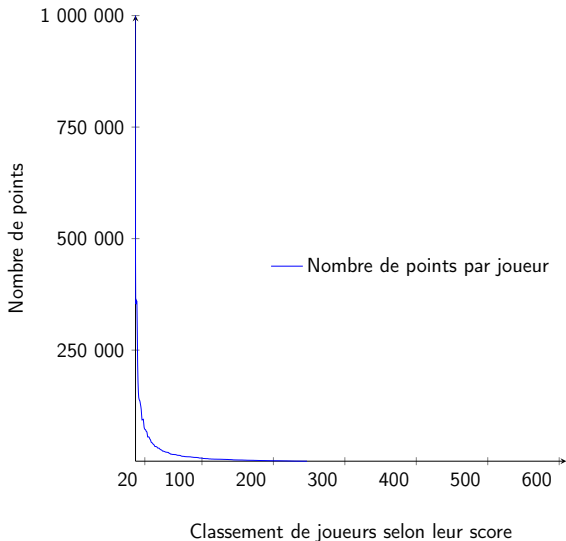
- ▶ plus de 23 000 en 1 semaine (semaine de la langue française)
- ▶ près de 50 000 en 1 mois

ZombiLingo : évaluation des annotations produites sur une référence multi-annotée et validée [Candito et al., 2014]

Login	Total ref. annotées	% correctes
<i>newK</i>	1 785	95,35%
<i>Pingu – kun</i>	985	55,74%
<i>xohwohxo</i>	909	88,78%
<i>JYA</i>	659	85,89%
<i>maelstromarie</i>	517	92,26%
<i>DihThi</i>	444	71,62%
<i>Zombinoclard</i>	382	77,49%
<i>bruno</i>	214	93,93%
<i>GG</i>	213	91,55%
<i>micHEL</i>	150	87,33%

- ▶ près de 84 % d'exactitude moyenne sur les 10 plus gros joueurs (plus de 86 % si on enlève le meilleur et le moins bon)
- ▶ plus de 90 % d'exactitude moyenne sur tous les joueurs

ZombiLingo : score (production) des joueurs



L'annotation par myriadisation par les jeux

- ▶ la qualité de la production des joueurs, quand elle est évaluable, est remarquable
- ▶ les quantités produites sont, elles, phénoménales
- ▶ relativement peu de participants produisent beaucoup et bien

... si les joueurs sont suffisamment bien formés (cas négatif des propriétés dans Phrase Detectives)

Une taxinomie (simplifiée) de la myriadisation

parmi d'autres, voir [Geiger et al., 2011]

Wikipedia, Projet Gutenberg :

- ▶ bénévole (non rémunérée)
- ▶ directe (le but à atteindre n'est pas masqué)

Games With A Purpose (Jeux ayant un but ou GWAP) :

- ▶ bénévole (non rémunérée)
- ▶ indirecte (le but à atteindre est plus ou moins masqué)

Amazon Mechanical Turk :

- ▶ rémunérée
- ▶ directe

Amazon Mechanical Turk et le travail parcellisé

MTurk

amazonmechanicalturk
Artificial Intelligence

Your Account

HITs

Qualifications

Already have an account?
Sign in as a [Worker](#) | [Requester](#)

[Introduction](#) | [Dashboard](#) | [Status](#) | [Account Settings](#)

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

179,373 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



Amazon Mechanical Turk et le travail parcellisé

MTurk est une plate-forme de **myriadisation** : le travail est *externalisé* via le Web et réalisé par de nombreuses personnes (la *foule*), ici les **Turkers**

The screenshot shows the Amazon Mechanical Turk website. At the top left is the logo "amazonmechanical turk" with the tagline "Artificial Intelligence". Navigation tabs include "Your Account", "HITS", and "Qualifications". A top right link says "Already have an account? Sign in as a Worker | Requester". Below the navigation is a header with "Introduction | Dashboard | Status | Account Settings". The main content area has a yellow banner with the text: "Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 179,373 HITS available. View them now." Below this are two columns. The left column is titled "Make Money by working on HITs" and describes HITs as individual tasks. It lists benefits for workers: "Can work from home", "Choose your own work hours", and "Get paid for doing good work". A flow diagram shows "Find an interesting task" (with a list of tasks like "Find terms and conditions", "Find supply chain", "Analyze enablement", "Turk TASKS analysis", "Government election", "Marketing. Globally", "See others") leading to "Work" (gears icon) and then "Earn money" (dollar sign icon). A "Find HITS Now" button is at the bottom. The right column is titled "Get Results from Mechanical Turk Workers" and describes asking workers to complete HITs. It lists benefits for requesters: "Have access to a global, on-demand, 24 x 7 workforce", "Get thousands of hits completed in minutes", and "Pay only when you're satisfied with the results". A flow diagram shows "Fund your account" (wallet icon), "Load your tasks" (gears icon), and "Get results" (star icon). A "Get Started" button is at the bottom.

amazonmechanical turk
Artificial Intelligence

Your Account | HITS | Qualifications

Already have an account? Sign in as a Worker | Requester

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.
179,373 HITS available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task **Work** **Earn money**

Find terms and conditions. Find supply chain. Analyze enablement. Turk TASKS analysis. Government election. Marketing. Globally. See others.

[Find HITS Now](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of hits completed in minutes
- Pay only when you're satisfied with the results

Fund your account **Load your tasks** **Get results**

[Get Started](#)

Amazon Mechanical Turk et le travail parcellisé

MTurk est une plate-forme de **myriadisation du travail parcellisé** : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée par les Requesters

The screenshot shows the Amazon Mechanical Turk website. At the top left is the logo "amazonmechanical turk" with "Artificial Intelligence" below it. Navigation tabs include "Your Account", "HITS", and "Qualifications". On the top right, it says "Already have an account? Sign in as a Worker | Requester". Below the navigation is a banner with the text: "Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 179,373 HITS available. View them now." The main content is split into two columns. The left column is titled "Make Money by working on HITS" and includes a green-bordered box with the text: "HITS - Human Intelligence Tasks - are individual tasks that you work on. Find HITS now." Below this, it says "As a Mechanical Turk Worker you:" and lists three bullet points: "Can work from home", "Choose your own work hours", and "Get paid for doing good work". A flow diagram shows "Find an interesting task" (with a globe icon), "Work" (with a gear icon), and "Earn money" (with a dollar sign icon), leading to a "Find HITS Now" button. The right column is titled "Get Results from Mechanical Turk Workers" and includes the text: "Ask workers to complete HITS - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now." Below this, it says "As a Mechanical Turk Requester you:" and lists three bullet points: "Have access to a global, on-demand, 24 x 7 workforce", "Get thousands of HITS completed in minutes", and "Pay only when you're satisfied with the results". A flow diagram shows "Fund your account" (with a wallet icon), "Load your tasks" (with a document icon), and "Get results" (with a star icon), leading to a "Get Started" button.

Amazon Mechanical Turk et le travail parcellisé

MTurk est une plate-forme de **myriadisation du travail parcellisé** : les tâches sont découpées en sous-tâches (HIT) et leur exécution est **payée**.

The screenshot shows the Amazon Mechanical Turk website. At the top left is the logo "amazonmechanical turk" with "Artificial Intelligence" below it. Navigation tabs include "Your Account", "HITS", and "Qualifications". A top right link says "Already have an account? Sign in as a Worker | Requester". Below the navigation is a banner with the text: "Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 179,373 HITS available. View them now." The main content is split into two columns. The left column is titled "Make Money by working on HITS" and describes HITS as Human Intelligence Tasks. It lists benefits for workers: working from home, choosing work hours, and getting paid for good work. A flow diagram shows "Find an interesting task" (with a globe icon), "Work" (with a gear icon), and "Earn money" (with a dollar sign icon), leading to a "Find HITS Now" button. The right column is titled "Get Results from Mechanical Turk Workers" and describes the process for requesters. It lists benefits: access to a global workforce, getting thousands of HITS quickly, and paying only for satisfactory results. A flow diagram shows "Fund your account" (with a plus sign icon), "Load your tasks" (with a document icon), and "Get results" (with a star icon), leading to a "Get Started" button.

amazonmechanical turk
Artificial Intelligence

Your Account | HITS | Qualifications

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.
179,373 HITS available. [View them now.](#)

Make Money by working on HITS

HITS - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITS now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → **Work** → **Earn money**

[Find HITS Now](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITS - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITS completed in minutes
- Pay only when you're satisfied with the results

Fund your account → **Load your tasks** → **Get results**

[Get Started](#)

Amazon Mechanical Turk et le travail parcellisé

MTurk est une plate-forme de **myriadisation du travail parcellisé** : les tâches sont découpées en sous-tâches (HIT) et leur exécution est **payée**.

The screenshot shows the Amazon Mechanical Turk website. At the top left is the logo "amazonmechanical turk" with the tagline "Artificial Intelligence". Navigation tabs include "Your Account", "HITS", and "Qualifications". A top right link says "Already have an account? Sign in as a Worker | Requester". Below the navigation is a banner with the text: "Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 179,373 HITS available. View them now."

The main content is divided into two columns:

- Left Column: Make Money by working on HITS**
 - Text: "HITS - Human Intelligence Tasks - are individual tasks that you work on. Find HITS now."
 - Section: "As a Mechanical Turk Worker you:"
 - Can work from home
 - Choose your own work hours
 - Get paid for doing good work
 - Flowchart: "Find an interesting task" (with a globe icon) → "Work" (with gear icons) → "Earn money" (with a dollar sign icon). A "Find HITS Now" button is at the bottom.
- Right Column: Get Results from Mechanical Turk Workers**
 - Text: "Ask workers to complete HITS - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now"
 - Section: "As a Mechanical Turk Requester you:"
 - Have access to a global, on-demand, 24 x 7 workforce
 - Get thousands of HITS completed in minutes
 - Pay only when you're satisfied with the results
 - Flowchart: "Fund your account" (with a plus sign icon) → "Load your tasks" (with a document icon) → "Get results" (with a star icon). A "Get Started" button is at the bottom.

Amazon Mechanical Turk : réalités

du travail à la tâche microrémunéré

Nombre de *Turkers* actifs [Fort et al., 2011] :

- ▶ nombre enregistré sur le site : plus de 500 000
 - ▶ 80 % des tâches (HIT) sont réalisées par les 20 % de *Turkers* les plus actifs [Deneme, 2009]
- ⇒ réellement actifs : entre 15 059 et 42 912

Qualité obtenue variable :

- ▶ insuffisante dès que la tâche est un peu complexe [Fort et al., 2011]
- ▶ le travail à la tâche microrémunéré (quelques cents) incite à produire en quantité, plutôt que de qualité

L'annotation manuelle de corpus, un domaine en friches

La question de l'expertise des annotateurs

La myriadisation, un point de vue unique sur l'annotation

Conclusion et perspectives

La myriadisation comme prisme de|sur l'annotation

- ▶ la **formation** des annotateurs est fondamentale [Dandapat et al., 2009, Bayerl and Paul, 2011] pour qu'ils puissent devenir des **experts de la tâche**
- ▶ il faut connaître la **complexité** de la tâche pour pouvoir la décomposer [Stubbs, 2012, Fort et al., 2012]
- ▶ l'**évaluation** doit avoir lieu tout au long de l'annotation [Bonneau-Maynard et al., 2005, Fort, 2012]

Myriadisation(s) et questions d'éthique

Éthique des jeux :

- ▶ prévention de l'addiction
- ▶ information aux joueurs : droits et devoirs, finalité du jeu, déclaration CNIL
- ▶ **ce qui est produit par la foule revient à la foule** : licence claire et disponibilité immédiate de la ressource créée

Myriadisation(s) et questions d'éthique

Éthique des jeux :

- ▶ prévention de l'addiction
- ▶ information aux joueurs : droits et devoirs, finalité du jeu, déclaration CNIL
- ▶ **ce qui est produit par la foule revient à la foule** : licence claire et disponibilité immédiate de la ressource créée

Éthique (minimale) du travail parcellisé (Amazon Mechanical Turk) :

- ▶ rémunération «décente»
- ▶ identification des travailleurs et des fournisseurs de tâches
- ▶ possibilité d'exclure un fournisseur de tâche pour «mauvaise conduite»

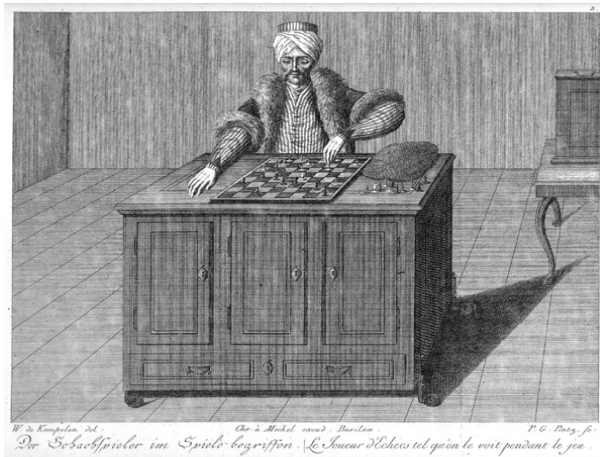
→ Un travail à la tâche peut-il être éthique ? (illégal en France)

Annexes

Amazon Mechanical Turk : compléments
Bibliographie

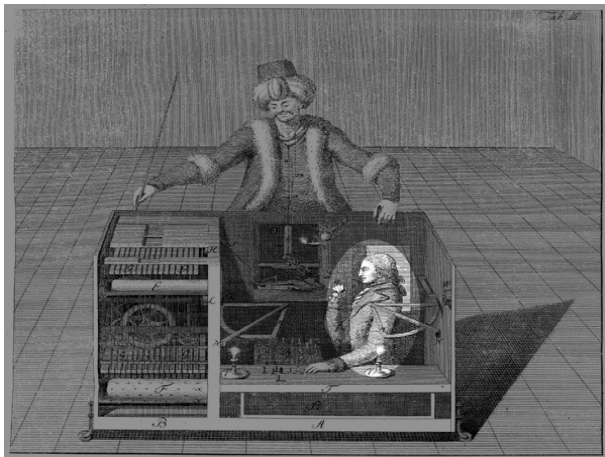
Le «Turc mécanique» de von Kempelen

Un joueur d'échecs mécanique créé par J. W. von Kempelen en 1770 :



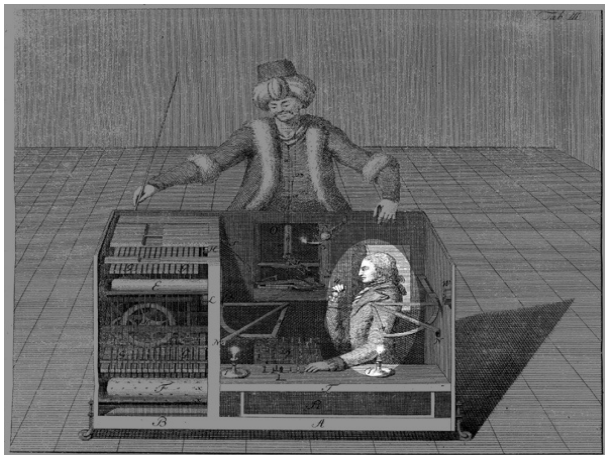
Le «Turc mécanique» de von Kempelen

En fait, un maître d'échecs était caché dans la machine :



Le «Turc mécanique» de von Kempelen

C'est l'intelligence artificielle **artificielle** !



Et Amazon créa Amazon Mechanical Turk

Amazon crée une pour ses propres besoins
plate-forme de travail parcellisé
et en ouvre l'accès en 2005 (moyennant 10 % des transactions)

Caractéristiques d'AMT

Rémunération :

- ▶ à la tâche (*illégal* en France sauf (rares) exceptions) : moins de 2 \$/h
- ▶ pas de relation explicite entre les *Turkers* et les *Requesters*

Tâches :

- ▶ nouveaux usages : par exemple, des créations artistiques, comme <http://www.thesheepmarket.com/>
- ▶ des tâches traditionnellement réalisées par des employés salariés : transcription, traduction (agences LDC, ELDA), etc

AMT : le rêve devenu réalité ?

Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks

Rion Snow[†] Brendan O'Connor[‡] Daniel Jurafsky[§] Andrew Y. Ng[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

[Snow et al., 2008]

AMT : le rêve devenu réalité ?

Cheap and Fast — But is it Good? **Evaluating Non-Expert Annotations for Natural Language Tasks**

Rion Snow[†] Brendan O'Connor[‡] Daniel Jurafsky[§] Andrew Y. Ng[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

[Snow et al., 2008]

C'est très peu cher, rapide, de bonne qualité
et c'est un hobby pour les *Turkers* !

AMT permet de réduire les coûts

Très basse rémunération \Rightarrow coûts faibles ? Oui, mais. . .

- ▶ coût de mise au point de l'**interface**
- ▶ coût de création de protections contre les **spammers**
- ▶ coût de **validation** et de **post-traitement**

certaines tâches (par exemple, la traduction du pachto vers l'anglais) génèrent des coûts similaires aux coûts habituels dans le domaine, du fait du **manque de Turkers qualifiés** [Novotney and Callison-Burch, 2010].

AMT permet de produire des ressources de qualité ?

- ▶ permet de produire des ressources de qualité dans certains cas précis (par exemple, la transcription simple)
- ▶ mais :
 - ▶ la qualité est insuffisante lorsque la tâche est **complexe** (par exemple, le résumé [Gillick and Liu, 2010])
 - ▶ l'**interface** d'AMT pose parfois problème [Tratz and Hovy, 2010]
 - ▶ les *Turkers* posent parfois problème (tricheurs, **spammers**)
 - ▶ le modèle de rémunération **à la tâche** pose problème [Kochhar et al., 2010]
- ▶ pour certaines tâches simples les outils de TAL produisent de **meilleurs résultats** qu'AMT [Wais et al., 2010].

AMT : un passe-temps pour les Turkers ?

[Ross et al., 2010, Ipeirotis, 2010] montre que :

- ▶ Turkers sont avant tout motivés par l'**argent** (91 %) :
 - ▶ 20 % considèrent AMT comme leur source de revenu primaire ;
 - ▶ 50 % comme leur source de revenu secondaire ;
 - ▶ l'aspect loisir n'est important que pour une minorité (30 %).
- ▶ 20 % des Turkers passent plus de 15 h par semaine sur AMT, et contribuent à 80 % des tâches.
- ▶ le salaire horaire moyen observé est **inférieur à 2 \$**.

Est-ce qu'AMT est éthique et/ou légal ?

Éthique :

- ▶ pas d'**identification** : pas de lien officiel entre *Requesters* et *Turkers* et entre *Turkers*
- ▶ pas de possibilité de **se syndiquer**, pour protester contre des manquements des *Requesters* ou ester en justice
- ▶ pas de **salaire minimum** (< 2 \$/h en moyenne)
- ▶ possibilité de **refuser de payer** les *Turkers*

Est-ce qu'AMT est éthique et/ou légal ?

The screenshot shows the Amazon Mechanical Turk homepage. At the top left is the logo "amazonmechanicalturk Artificial Intelligence". Navigation tabs include "Your Account", "HITS", and "Qualifications". A top right link says "Already have an account? Sign in as a Worker | Requester". A central banner reads "Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 179,373 HITS available. View them now." Below this are two main sections: "Make Money by working on HITS" and "Get Results from Mechanical Turk Workers".

Make Money by working on HITS

HITS - Human Intelligence Tasks - are individual tasks that you work on. [Find HITS now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

The process flow is: **Find an interesting task** (represented by a gear icon with text: "Find an interesting task", "Find tasks, and supply them to Mechanical Turk workers to complete your tasks") → **Work** (represented by a gear icon) → **Earn money** (represented by a dollar sign icon). A button "Find HITS Now" is at the bottom.

Get Results from Mechanical Turk Workers

Ask workers to complete HITS - Human Intelligence Tasks - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITS completed in minutes
- Pay only when you're satisfied with the results

The process flow is: **Fund your account** (represented by a plus sign icon) → **Load your tasks** (represented by a gear icon with a plus sign) → **Get results** (represented by a star icon). A button "Get Started" is at the bottom.

Est-ce qu'AMT est éthique et/ou légal ?

Légalité :

- ▶ accord de licence d'Amazon : les *Turkers* sont considérés comme des travailleurs indépendants ⇒ ils sont supposés se déclarer comme tels et payer les cotisations afférentes
- ▶ illusoire, vu le niveau de rémunération
- ⇒ les États **perdent** une source de revenus légitime



Bayerl, P. S. and Paul, K. I. (2011).

What determines inter-coder agreement in manual annotations? a meta-analytic investigation.

Computational Linguistics, 37(4) :699–725.



Böhmová, A., Hajič, J., Hajičová, E., and Hladká, B. (2001).

The prague dependency treebank : Three-level annotation scenario.

In Abeillé, A., editor, Treebanks : Building and Using Syntactically Annotated Corpora. Kluwer Academic Publishers.



Bonneau-Maynard, H., Rosset, S., Ayache, C., Kuhn, A., and Mostefa, D. (2005).

Semantic annotation of the French Media dialog corpus.

In Proceedings of the InterSpeech, Lisbonne, Portugal.



Candito, M., Perrier, G., Guillaume, B., Ribeyre, C., Fort, K., Seddah, D., and de la Clergerie, E. (2014).

Deep syntax annotation of the sequoia french treebank.

In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Islande.



Candito, M. and Seddah, D. (2012).

Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical.

In Actes de Traitement Automatique des Langues Naturelles (TALN), Grenoble, France.



Chamberlain, J., Kruschwitz, U., and Poesio, M. (2009).

Constructing an anaphorically annotated corpus with non-experts : assessing the quality of collaborative annotations.

In Proceedings of the 2009 Workshop on The People's Web Meets NLP : Collaboratively Constructed Semantic Resources, People's Web '09, pages 57–62, Stroudsburg, PA, USA. Association for Computational Linguistics.



Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008).

Phrase Detectives : a web-based collaborative annotation game.

In Proceedings of the International Conference on Semantic Systems (I-Semantics'08), Graz, Autriche.



Dandapat, S., Biswas, P., Choudhury, M., and Bali, K. (2009).
Complex linguistic annotation - no easy way out! a case from
bangla and hindi POS labeling tasks.

In Proceedings of the third ACL Linguistic Annotation
Workshop, Singapour.



Deneme (2009).

How many turkers are there?

<http://groups.csail.mit.edu/uid/deneme/>.



Fort, K. (2012).

Les ressources annotées, un enjeu pour l'analyse de contenu :
vers une méthodologie de l'annotation manuelle de corpus.

PhD thesis, Université Paris XIII, LIPN, INIST-CNRS.



Fort, K., Adda, G., and Cohen, K. B. (2011).

Amazon Mechanical Turk : Gold mine or coal mine ?

Computational Linguistics (editorial), 37(2) :413-420.



Fort, K., Guillaume, B., and Chastant, H. (2014).

Creating Zombilingo, a Game With A Purpose for dependency syntax annotation.

In

Gamification for Information Retrieval (GamifIR'14) Workshop, Amsterdam, Pays-Bas.



Fort, K., Nazarenko, A., and Rosset, S. (2012).

Modeling the complexity of manual annotation tasks : a grid of analysis.

In Proceedings of the International Conference on Computational Linguistics (COLING), pages 895–910, Mumbai, Inde.



Fort, K. and Sagot, B. (2010).

Influence of pre-annotation on POS-tagged corpus development.

In Proceedings of the Fourth ACL Linguistic Annotation Workshop, pages 56–63, Uppsala, Suède.



Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., and Schader, M. (2011).

Managing the crowd : Towards a taxonomy of crowdsourcing processes.

In AMCIS 2011 Proceedings.



Gillick, D. and Liu, Y. (2010).

Non-expert evaluation of summarization systems is risky.

In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT '10, pages 148–151, Stroudsburg, PA, USA. Association for Computational Linguistics.



Howe, J. (2006).

The rise of crowdsourcing.

Wired Magazine, 14(6).



Ipeirotis, P. (2010).

The new demographics of mechanical turk.

<http://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html>.



Kim, J.-D., Ohta, T., and Tsujii, J. (2008).

Corpus annotation for mining biomedical events from literature.

BMC Bioinformatics, 9(1) :10.



Kochhar, S., Mazzocchi, S., and Paritosh, P. (2010).

The anatomy of a large-scale human computation engine.

In Proceedings of Human Computation Workshop at the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2010, Washington D.C.



Mathet, Y., Widlöcher, A., Fort, K., François, C., Galibert, O., Grouin, C., Kahn, J., Rosset, S., and Zweigenbaum, P. (2012).

Manual corpus annotation : Evaluating the evaluation metrics.

In Proceedings of the International Conference on Computational Linguistics (COLING), pages 809–818, Mumbai, Inde.

Poster.



Novotney, S. and Callison-Burch, C. (2010).

Cheap, fast and good enough : automatic speech recognition with non-expert transcription.

In Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), HLT'10, pages 207–215, Stroudsburg, PA, USA. Association for Computational Linguistics.



Péry-Woodley, M.-P., Afantenos, S., Ho-Dac, L.-M., and Asher, N. (2011).

La ressource ANNODIS, un corpus enrichi d'annotations discursives.

Traitement Automatique des Langues, 52(3) :71–101.



Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. (2010).

Who are the crowdworkers? : shifting demographics in mechanical turk.

In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, CHI EA '10, pages 2863–2872, New York, NY, USA. ACM.



Sagot, B., Fort, K., Adda, G., Mariani, J., and Lang, B. (2011).

Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé.

In Actes de Traitement Automatique des Langues Naturelles (TALN), Montpellier, France.

12 pages.



Snow, R., O'Connor, B., Jurafsky, D., and Ng., A. Y. (2008). Cheap and fast - but is it good? evaluating non-expert annotations for natural language tasks.

In Proceedings of EMNLP 2008, pages 254–263.



Stubbs, A. (2012).

Developing specifications for light annotation tasks in the biomedical domain.

In Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining, Istanbul, Turkey.



Tellier, M. (2014).

Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés. Discours, 15.



Tratz, S. and Hovy, E. (2010).

A taxonomy, dataset, and classifier for automatic noun compound interpretation.

In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 678–687, Uppsala, Suède. Association for Computational Linguistics.



Wais, P., Lingamneni, S., Cook, D., Fennell, J., Goldenberg, B., Lubarov, D., Marin, D., and Simons, H. (2010).

Towards building a high-quality workforce with mechanical turk. In Proceedings of Computational Social Science and the Wisdom of Crowds (NIPS).